Instagram Posts Data Analysis

Name: SWATI SHARMA

OBJECTIVE:

- The objective of this report is to analyze engagement metrics of Instagram posts using a dataset containing key attributes such as Post ID, Post Type, number of Comments, and number of Likes.
- Through data analysis conducted in R and visualized via a Power BI dashboard, the goal is to uncover patterns and insights related to post performance.
- This includes identifying which types of posts drive higher engagement, understanding user interaction behavior, and supporting data-driven decisions for optimizing future Instagram content strategy.

Dataset Description:

- Source: Kaggle (Instagram Posts Dataset)
- Rows: 4 records
- Key columns: Post_id, Post_Type, comments, likes

Tools & Technologies Used:

- **Excel**
- **R**
- **Power BI**

Code and Output of Python:

```
install.packages("ggplot2")
```

```
> install.packages("ggplot2")
```

install.packages("dplyr")

```
> install.packages("dplyr")
```

#Loading the required libraries

library(ggplot2)

library(dplyr)

```
> #Loading the required libraries
> library(ggplot2)
> library(dplyr)
```

##open a file browser window where you can manually select your file

data <- read.csv(file.choose(), stringsAsFactors = FALSE)

# View the first few rows
head(data)

```
  Post_id Post_Type comments likes
1     101  Carousel      268 16382
2     102      Reel      138  9267
3     103      Reel     1089 10100
4     104      Reel      271  6943
5     105      Reel      145 17158
6     106      Reel      143  9683
>
```

# Structure of the data
str(data)

```
'data.frame':   35 obs. of  4 variables:
 $ Post_id  : int  101 102 103 104 105 106 107 108 109 110 ...
 $ Post_Type: chr  "Carousel" "Reel" "Reel" "Reel" ...
 $ comments : int  268 138 1089 271 145 143 132 128 884 98 ...
 $ likes    : int  16382 9267 10100 6943 17158 9683 4287 7484 48528 6754
...
```

Q1) How many total posts are there?

```
#Total number of posts
nrow(data)
```

```
[1] 35
```

Q2) What are the different types of posts?

```
#Different types of posts

unique(data$Post_Type)
```

```
[1] "Carousel" "Reel"     "Image"
```

Q3) How many posts are there for each post type?

```
#Number of posts for each post type

table(data$Post_Type)
```

```
Carousel     Image      Reel
      14         9        12
```

Q4) What is the average number of likes across all posts?

```
#Average likes overall

mean(data$likes, na.rm = TRUE)
```

```
[1] 19797.29
```

Q5) What is the average number of comments across all posts?

```
#Average comments overall

mean(data$comments, na.rm = TRUE)
```

```
[1] 223.8286
>
```

Q6) Which post type receives the most engagement (likes + comments)?

```
#Average engagement by post type

data$engagement <- data$likes + data$comments

aggregate(engagement ~ Post_Type, data = data, FUN = mean)
```

```
  Post_Type engagement
1  Carousel   20988.57
2     Image   22744.56
3      Reel   16849.83
>
```

Q7) What is the total number of likes and comments for each post type?

```
#Total likes and comments by post type

data %>%

 group_by(Post_Type) %>%

 summarise(Total_Likes = sum(likes, na.rm = TRUE),

           Total_Comments = sum(comments, na.rm = TRUE))
```

```
# A tibble: 3 × 3
  Post_Type Total_Likes Total_Comments
  <chr>           <int>          <int>
1 Carousel       291400           2440
2 Image          203199           1502
3 Reel           198306           3892
>
```

Q8) Which post has the highest number of likes?

```
#Post with the highest likes

data[which.max(data$likes), ]
```

```
   Post_id Post_Type comments likes engagement
25     125  Carousel      466 79000      79466
>
```

Q9) Which post has the highest number of comments?

```
#Post with the highest comments

data[which.max(data$comments), ]
```

```
  Post_id Post_Type comments likes engagement
3     103      Reel     1089 10100      11189
> |
```

Q10) Which post has the lowest engagement?

```
#Post with the lowest engagement

data[which.min(data$engagement), ]
```

```
   Post_id Post_Type comments likes engagement
11     111     Image        1   160        161
> |
```

Q11) What is the distribution of likes and comments for each post type?

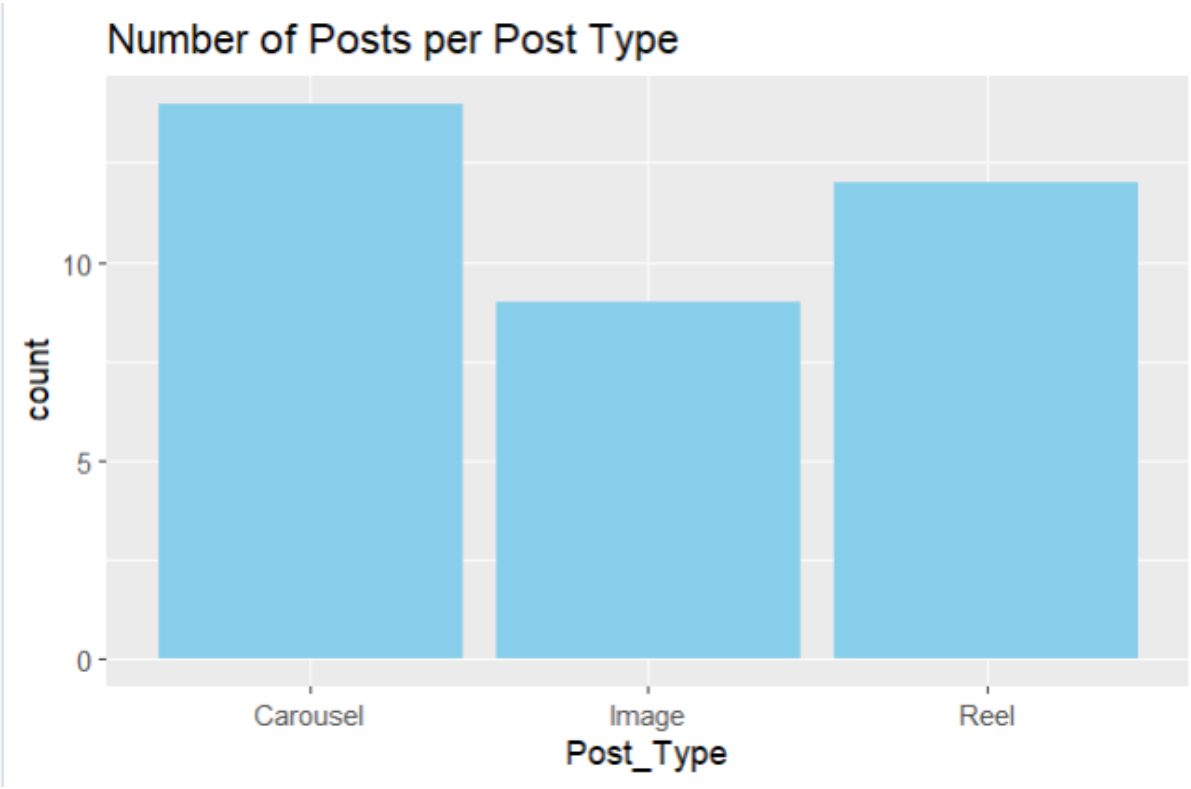
```
#Distribution of likes/comments by post type (summary stats)

data %>%
  group_by(Post_Type) %>%
  summarise(Mean_Likes = mean(likes, na.rm = TRUE),
            SD_Likes = sd(likes, na.rm = TRUE),
            Mean_Comments = mean(comments, na.rm = TRUE),
            SD_Comments = sd(comments, na.rm = TRUE))
```

```
# A tibble: 3 × 5
  Post_Type Mean_Likes SD_Likes Mean_Comments SD_Comments
  <chr>          <dbl>    <dbl>         <dbl>       <dbl>
1 Carousel      20814.   26888.          174.        190.
2 Image         22578.   24272.          167.        165.
3 Reel          16526.   16481.          324.        339.
> |
```

Visualization

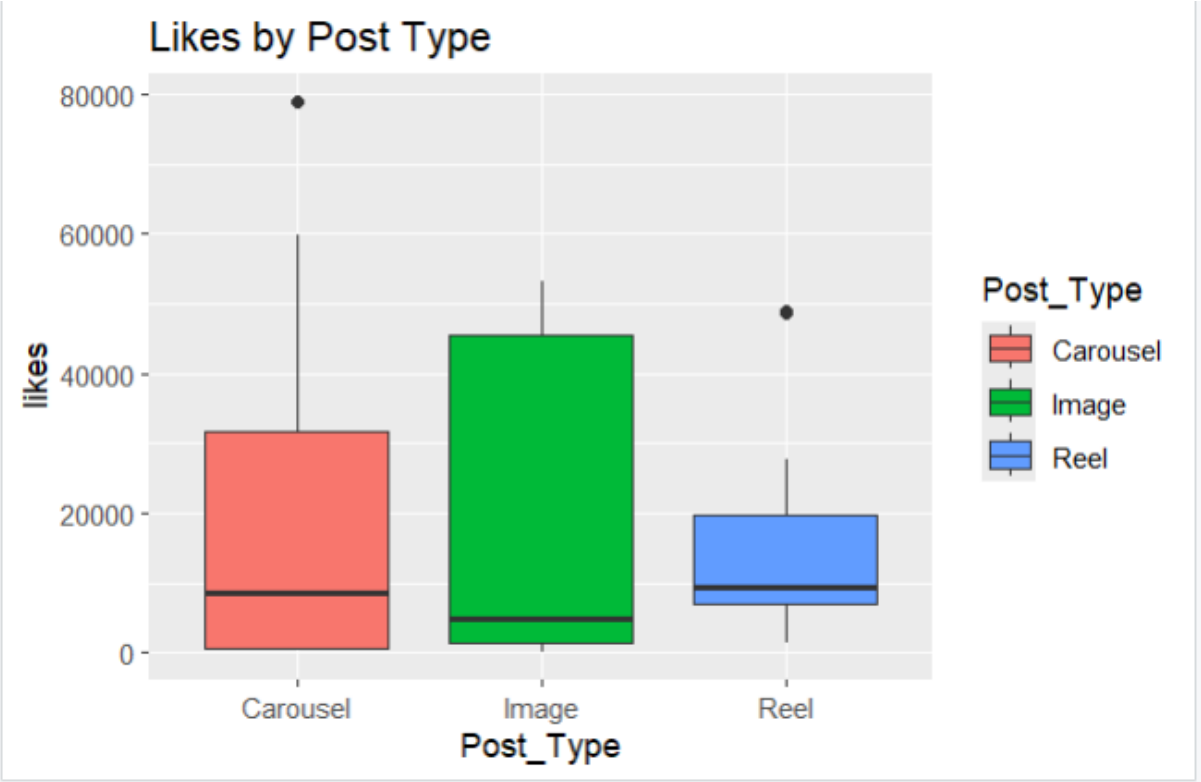
Q12) Bar chart of total posts per post type?

```
#Bar chart: total posts per post type
ggplot(data, aes(x = Post_Type)) +
  geom_bar(fill = "skyblue") +
  labs(title = "Number of Posts per Post Type")
```



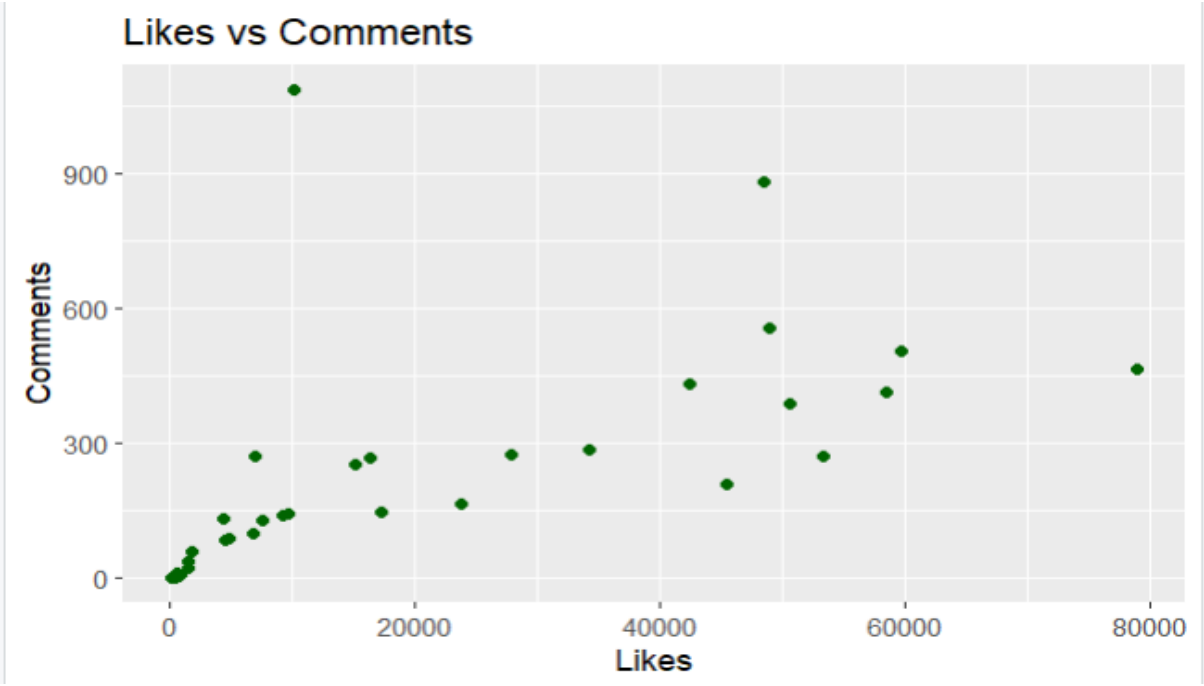
Q13) Boxplot of likes by post type?

```
#Boxplot of likes by post type
ggplot(data, aes(x = Post_Type, y = likes, fill = Post_Type)) +
  geom_boxplot() +
  labs(title = "Likes by Post Type")
```



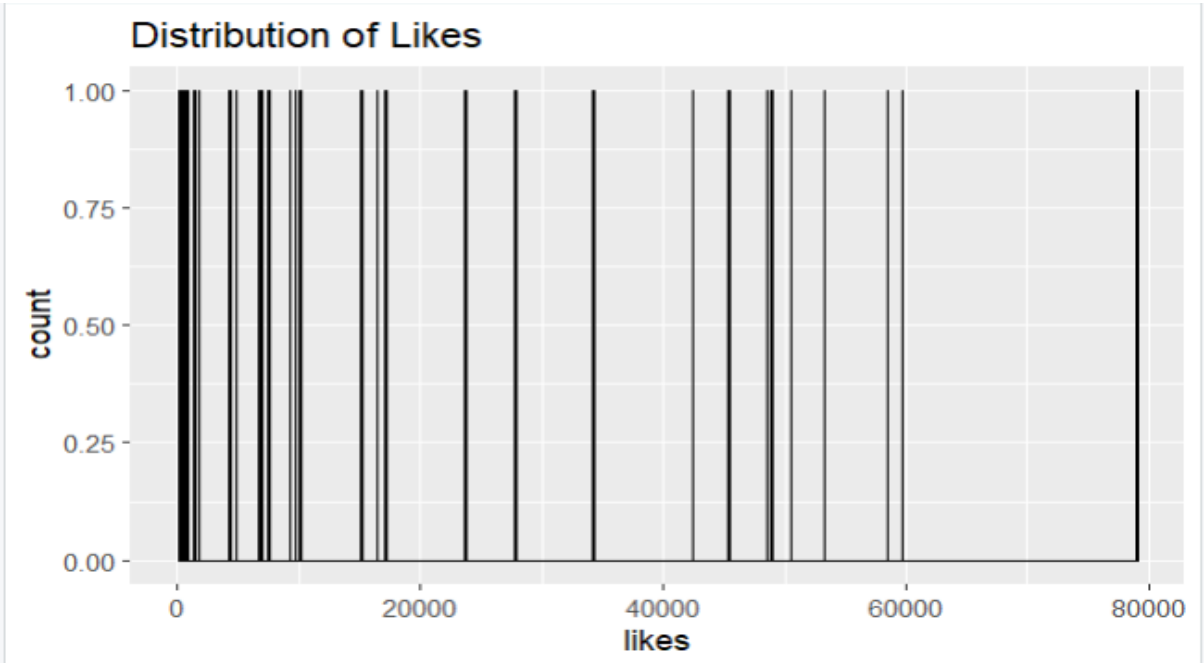
Q14) Scatter plot of likes vs comments?

```
#Scatter plot: likes vs comments
ggplot(data, aes(x = likes, y = comments)) +
  geom_point(color = "darkgreen") +
  labs(title = "Likes vs Comments", x = "Likes", y = "Comments")
```

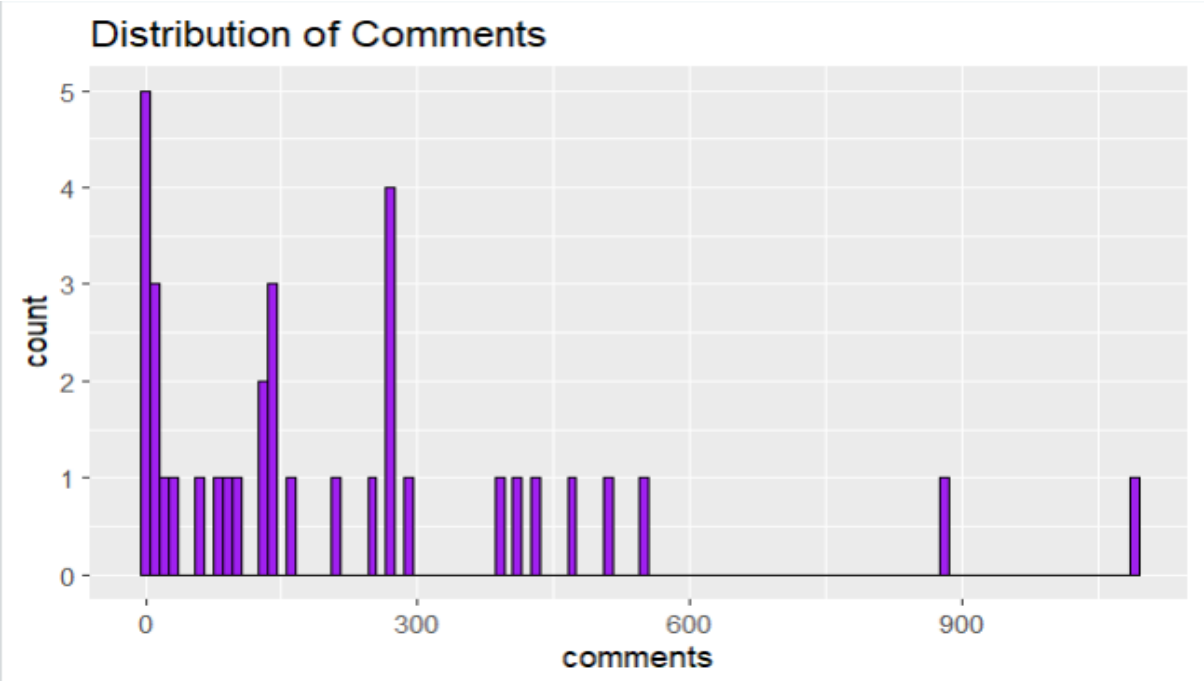


Q15) Histogram of likes or comments to see their distribution?

```
#Histogram of likes
ggplot(data, aes(x = likes)) +
  geom_histogram(binwidth = 50, fill = "orange", color = "black") +
  labs(title = "Distribution of Likes")
```



```
# Histogram of comments
ggplot(data, aes(x = comments)) +
  geom_histogram(binwidth = 10, fill = "purple", color = "black") +
  labs(title = "Distribution of Comments")
```



Q16) Is there a correlation between likes and comments?

```
#Correlation between likes and comments
cor(data$likes, data$comments, use = "complete.obs")
```

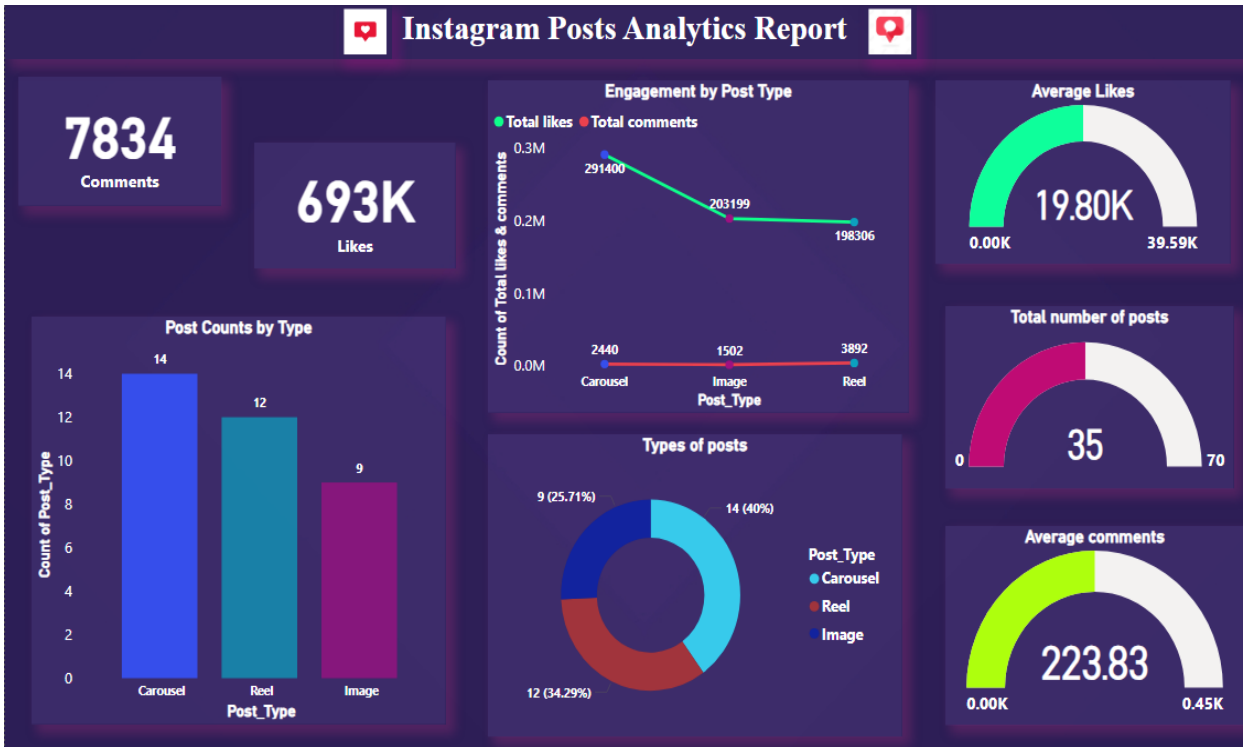
```
[1] 0.6173911
>
```

Q17) Are there outliers in the number of likes or comments?

```
#Outliers: likes > 99th percentile
likes_threshold <- quantile(data$likes, 0.99, na.rm = TRUE)
outliers <- data %>% filter(likes > likes_threshold)
outliers
```

```
  Post_id Post_Type comments likes engagement
1     125  Carousel      466 79000      79466
> 
```

Screenshot of Dashboard:



Method:

❖ Data Collection

- Downloaded Instagram Posts dataset from Kaggle.
- Dataset includes various features such as Player Post_id, Post_Type, comments, likes

❖ Data Cleaning & Preprocessing

- Excel: I used **Microsoft Excel** for data cleaning.
- Removed duplicates(Filling, removing, or imputing null/blank data)
- Handled missing values
- Filtering irrelevant Data (removing unnecessary columns and rows)

❖ Data Loading to R

- Before loading, basic checks and cleaning operations were performed.

- Before loading the data, necessary packages were installed and the R environment was prepared.
- Required libraries were loaded.
- The dataset was imported in CSV format using an interactive file selection method.
- With the cleaned dataset, further analysis was carried out to extract insights and generate visualizations.

❖ **Pattern Identification**

- Total number of likes.
- Total number of comments.
- Average number of likes.
- Average number of comments.
- Total number of posts.
- Post count by types.
- Types of post.
- Engagement by post types.

❖ **Key Insights**

- Total Number of Likes: 693,000
- Total Number of Comments: 7,834
- Average Number of Likes: 19,800
- Average Number of Comments: 223.83
- Total Number of Posts: 35
- Post Type Distribution

Post Type Number of Posts

Carousel 14

Reels 12

Images 9

Types of Posts Included in the Dataset:

- Carousel
- Reels
- Images

- Engagement by Post Types

Post Type	Total Likes	Total Comments
Carousel	291,400	2,440
Images	203,199	1,502
Reels	198,306	3,892

❖ **Reporting & Visualization**

➢ Tools Used: Power BI

❖ **Conclusion**

- **Informs Content Strategy:** Understanding which post types generate the most likes and comments enables targeted content creation to maximize engagement.
- **Optimizes Resource Allocation:** Focus efforts and budget on post formats that deliver the best interaction returns.
- **Enhances Audience Engagement:** Tailoring posts based on engagement patterns helps build stronger relationships with followers.
- **Supports Growth:** By leveraging insights on post-performance, marketers can increase reach, boost brand awareness, and drive conversions.